Predicting the Competency Improvement for Serious Games Analytics: Action-sequences, Game Grids, PLS DA and JMP

Christian Sebastian Loh & I-Hung Li
Virtual Environment Lab (V-Lab)
Southern Illinois University
Carbondale, IL

Abstract

Serious games were originally meant to become advanced training tools to improve decision-making skills and raise job performance in trainees/learners. Currently, less than 10% of serious games have been designed to facilitate training – including those used for military and medical simulation and training. Serious games analytics can increase the value of these types of training if we know what performance metrics to use, how to turn these analytics to meaningful insights for stakeholders, as well as provide just-in-time (re)training and remediation for the learners, which can impact Return of Investment (ROI) for the learning organizations.

The behavioral and cognitive differences between skilled individuals (experts) and novices have been well-documented in the literature. Action sequence, a component in competency, comprised of the chronological order of actions performed, has been shown to be useful in differentiating (likely) experts from novices for training performance assessment purposes. The ability to discriminate experts from novices and possibly to predict who might become experts faster (shorter period of training) is highly desirable in the training and learning industries.

We captured the actions of players in a serious games and coded their action-sequences by dividing the game world into grids of different granularity. We use the Partial Least Squares Discriminant Analysis (PLS DA) function in JMP Pro 12 to compare these variables, along with total game completion time, to differentiate the players into experts and novices based on their behaviors in the game for performance assessment and *serious games analytics*.

Keywords: Partial Least Squares Discriminant Analysis, Serious Games Analytics, action sequences, similarity index, performance assessment

Introduction

When serious games were first commissioned for military training in the U.S. (Krulak, 1997), they were meant to become advanced training tools to improve decision-making skills and raise job performance in the learners/trainees. Unfortunately, about 90% of ‘serious games’ produced turned out to be *message broadcasters* (Alvarez, Djaouti, Rampnoux, & Alvarez, 2011) – i.e., well-design ‘messages’ for dissemination purposes –

true to their communications, advertising, and art/design backgrounds, but lacking in assessment needed to verify training efficacy. It is no wonder that an overwhelming majority of training and educative 'serious' games remain weak in performance assessment and primarily focus on educative message broadcasting. Of the remaining 10% of the 'serious games' that had been geared towards scenario-based training – including those for military and medical training/simulation, many of them also lack assessment.

Despite many calls for serious games and simulations to add more value by incorporating appropriate debriefing tools that evaluate paly-learners' performance on-the-fly for performance assessment and improvement (Crookall, 2010), progress in this area has been slow. Given the diverse requirements among researchers (seeking improvement in methods), serious games developers (seeking profits) and stakeholders of learning organizations (seeking performance improvement), many (re)alignments must be done to put into serious games the assessment and feedback mechanisms needed to generate just-in-time (re)training and remediation analytics: Serious games analytics need to be useful – facilitate the calculation of Return of Investment (ROI) for learning organizations (Loh, 2012).

Towards this end, it would do well for all stakeholders to look into analytical methods involving statistical/machine learning, data analytics, and visualization to better trace and analyze user-behaviors *in situ* within the game-based learning habitats, which can be used to justify cost and improve profits (Loh, 2011). Given that serious games of today can increasingly accrue big (online) data and produce *serious games analytics* (Loh, Sheng, & Ifenthaler, 2015), one needs to take advantage of advance statistical methods to better analyze the gameplay data via (un)supervised learning for analytics and visualizations. Compare this approach to the current prevalent methods of surveys and pretest/posttests (Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013), which treats serious games as impenetrable Black Boxes (Loh & Sheng, 2015).

Transforming Behavioral Actions into Coded Sequences

If serious games can be likened to the ancient Gordian knot, then expertise would be the sword that slices through the mass. Expertise research is interested in knowing how individuals improve their performance in executing task(s) through (deliberate) practice (Ericsson, Prietula, & Cokely, 2007). Differences between experts and novices has been demonstrated in many areas, including time-to-task-completion (Cappiello et al., 2011; Hornbæk, 2006), mental representations (Kozma & Russell, 1997), gaze patterns (Law, Atkins, Kirkpatrick, & Lomax, 2004), neural/perceptual responses (Mishra, Zinni, Bavelier, & Hillyard, 2011), and digital gameplay (Loh & Sheng, 2013).

A common first step in measuring the behavioral-action differences in individuals is to convert their action traces into a series of numerical sequences, called *action sequences*. Once in numerical form, these action sequences can then be analyzed using statistical software, such as JMP, SAS, or R. Naturally, inappropriately coded action-sequences often lead to GIGO (Garbage In, Garbage Out) that impede subsequent steps in the serious games analytics process.

In the following sections, we will describe: (a) how we encode the actions of two groups of play-learners (experts vs. novices) into action sequences, (b) how we use the action sequences encoded to help us predict the expertise category of an unknown group of play-learners, and how we used the Partial Least Squares Discriminant Analysis (PLS DA) function in JMP Pro 12 to visualize the data as analytics. This report constitutes some of the serious games analytics research (Loh, Sheng, & Ifenthaler, 2015) currently being pioneered at the Virtual Environment Lab (V-Lab).

From Action Sequences to Similarity Index

Using an in-house developed serious game, we telemetrically captured user-generated data with *Information Trails* (see Loh, Anantachai, Byun, & Lenox, 2007; Loh, 2012b). We divided the game world into 10 (square) grids of various granularity, from very coarse to very fine – i.e., 5×5, 10×10, 15×15, 20×20, 25×25, 30×30, 35×35, 40×40, 45×45, and 50×50. Using the grids as templates, we proceeded to transform all players' in-game movement into action sequences.

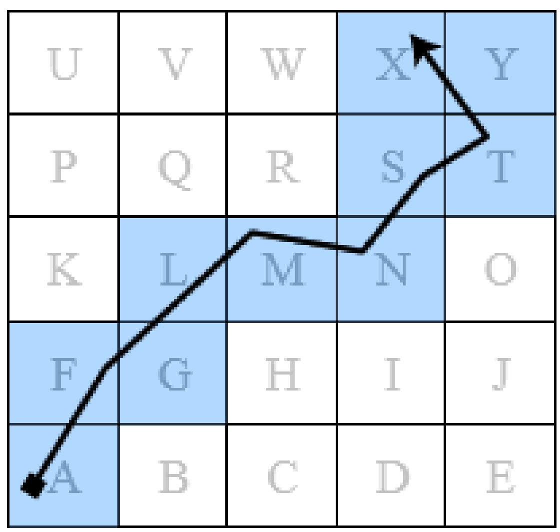

Figure 1. The navigational path of a player over a 5×5 game grid.

Figure 1 above depicts a sample game world that has been divided into a 5×5 grid. The action sequence {AFGLMNSTYX} represents the player's movement over the game grid – shown as a directional black line. Naturally, as the size of the game grid became smaller, the action sequences will become much longer.

In order to measure the (dis)similarities between the action sequences of novices versus that of the experts, we calculated their corresponding similarity indices – specifically, the Jaccard coefficient – against one *single* expert, using the method depicted in Loh & Sheng (2015b). The value of the Jaccard coefficient is given by the following formula:

$$Jaccard\ (a,b) = \frac{|a \cap b|}{|a \cup b|}, where\ 0\ \leq Jaccard \leq 1$$

Instructions to code the action sequence into bigrams for incorporation into the above formula is available in Loh & Sheng (2013). The method to calculate similarity indices in the presence of multiple experts (i.e., Maximum Similarity Indices or MSI) is available in Loh & Sheng (2014).

Area Revisitation

In calculating similarity indices, the repeated sections of a string are systematically removed to reduce the length of analyzed sequences. This approach suited the intent of Record Linkage analysis (Winkler, 2006) to clean out large databases through the removal of extraneous or duplicated data for data mining (Monge & Elkan, 1997). Player *a*, who traversed the game world with an action sequence of {ABCBCBCBCBCBCD} would ended up having the same Jaccard coefficient with player *b*, with an action sequence of {ABCBCD}. The calculation is shown below:

$$Jaccard\ (a,b) = \frac{|a \cap b|}{|a \cup b|} = \frac{[AB,BC,CB,CD]}{[AB,BC,CB,CD]} = \frac{4}{4} = 1$$

In (serious) games research, using the same approach to calculate action sequences may severely impact the resulting analytics because players are known to revisit a game area for a variety of reasons (see Thawonmas, Yoshida, Lou, & Chen, 2009, 2011). Besides becoming ‘lost’ in a massive game world, other legitimate reasons for game players to revisit an area may include: mining, grinding, or simply finding out if non-player characters have new information to share. Moreover, a player who has become lost – i.e., kept going round and round (like player *a*) in a game world should not be regarded to have the same performance or skills with a player who could complete the game in fewer moves or turns (e.g., player *b*).

Thus, we decided to code the action sequences of revisitation differently by regarding the revisitation of a game grid as a new/unique event (e.g., STSXS $\rightarrow$ $S_1TS_2XS_3$). This deliberate step was performed to prevent over-simplification of players’ in-game movement, which may erroneously equate a novice as an expert (shown in the example above). While this may ‘take away’ some features of the similarity metric, we believe it makes sense within the domain knowledge and is necessary for the compensation of area revisitation – an integral feature forming unique gameplay pattern in (serious) games.

Control Group

While many researchers (both industrial and educational) continue to compare a game-playing group against a second non-game playing group (as control) in game research, this notorious and flawed “Media Comparison” design was debunked in 1980s (see Clark, 1985, 1992, 1994; Hastings & Tracey, 2004), and should be stopped at all costs.

We hold that there are other better alternative research design for serious game research than comparing them to a control group. Since we are comparing the similarities of expert-novice action sequences obtained using the same training environment, the control group method is not applicable in this case.

Materials and Methods

The serious game narrative in this study comprised of two different *ground reconnaissance* tasks with of six small objectives. The number of objectives were determined based on the total amount of game time (about 1-2 hours) that gave us a reasonable amount of player-generated data, without the risk of causing player fatigue. Since total game time would be an important factor related to the movement within a game world, we included *time* as a predictor.

To begin the game, players must login to the game server, which capture their login ID and sets up the appropriate user database. Upon entering the game world, a mission giver would immediately engage and inform the players of two ground reconnaissance tasks with six mini objectives. After this, a nearby gate would be unlocked to release the players into the actual game area. Give the nonlinear game narrative, players were free to explore the game world and complete the objectives in any order they chose. Once all the tasks have been completed, the players must speak with the mission giver once again to end the game.

A total of 62 players from a mid-Western public university took part in the study. Consent to collect data was obtained meaning the participants had full knowledge that their gameplay data would be recorded. There was no indication to suggest this knowledge affected their game playing in any way. The players generated *in situ* 534,837 raw (gameplay) data points. Of all 62 players, 7 of them were known experts. The rest (55) were designated as novices with unverified performance.

We created a *Performance Tracing Report Assistant* (PeTRA) to automate (a) the coding of navigational paths into action sequences (depending on the granularity of the grids), (b) the calculations of Jaccard coefficients from the action sequences for all 10 grids (from $J_{5\times5}$, $J_{10\times10}$, … to $J_{50\times50}$).

A curvilinear relationship was detected between the *time* variable and the Jaccard coefficients, and log transformation of time was performed, resulting in *log time* as the new predictor variable. We coded the *expert* group using a binary dummy variable (0 for novice and 1 for expert) and obtained the simple bivariate correlations to evaluate the linear relationships between each of the aforementioned *Jaccard* variables (10) and *log time* (shown as Table 1).

Table 1. Bivariate correlations between Expert and the Jaccard coefficients (n = 62, all significant at $p < .01$)

	$J_{5\times5}$	$J_{10\times10}$	$J_{15\times15}$	$J_{20\times20}$	$J_{25\times25}$	$J_{30\times30}$	$J_{35\times35}$	$J_{40\times40}$	$J_{45\times45}$	$J_{50\times50}$	log time
Expert	.586	.692	.579	.480	.430	.418	.410	.368	.406	.375	-.655

Data Analysis

As the advances of technologies brought about more and more data, new statistical methods are being developed to help analyze high-dimensional data often found in the field of statistics, machine learning, biology, and bioinformatics (Boulesteix & Strimmer, 2007). More recently, such method has been also been extended to chemistry (Ballabio & Consonni, 2013) and serious games analytics (Loh, Sheng, & Li, 2015).

Partial Least Squares Discriminant Analysis (PLS DA) is a linear classification method involving part PLS regression and part DA (Ballabio & Consonni, 2013). PLS DA is very effective when there are more predictors than observations, a situation in which other models would not work.

PLS DA and JMP Pro 12

By combining the dimensional reduction ability of PLS regression (Boulesteix & Strimmer, 2007) and the supervised pattern recognition (i.e., classification) ability of DA, PLS DA is able to predict group membership of unknown observations based on features of the known *a priori* groups where many traditional method are not applicable. Given that the number of predictors exceeds the number of observations in this study, PLS DA naturally became the analytical method of choice.

PLS DA is now available in JMP Pro 12, accessed via the ***Fit Model*** option and assigning a nominal variable as ***Y***. Following the guidelines suggested by Cox & Gaudard (2013), we:

1. Entered the nominal variable, *Effect*, as ***Y***.
2. Entered all *predictors* (i.e., 11 *Jaccard coefficients and log time*) as ***Model Effects***.
3. Selected ***Partial Least Squares*** as the ***Personality***.
4. Deselected the ***Standardize X*** option, and clicked ***Run***.
5. Selected ***NIPALS*** as the ***Method Specification***.
6. Selected ***k-Fold*** (with 7 folds) as the ***Validation Method***, and clicked ***Go***.

The process fitted a 2-factor model with a minimum root mean PRESS of .8524 that is capable of explaining 97.748% of total variance for Cumulative X and 46.924% for Cumulative Y. The percentage of variation explainable by each factor is shown in Table 2.

Table 2: k-Fold Cross Validation with K = 7 and Method = NIPALS, and Percent of Variation Explained

No. of Factors	Root Mean PRESS	Van der Voet T^2	Q^2	X Effects	% Variation Explained X	Y Responses	% Variation Explained Y
0	1.0722	3.476	-.0003				
1	4.2311	1.012	-14.838	79.898	79.898	34.431	34.431
2	.8524	.000	.368	17.850	97.748	12.493	46.924

Data Visualizations

The Variables Importance Plot (VIP) showed all variables to have a VIP value exceeding .80 (Figure 2) – meaning that these are all important for the modeling of both **X** and **Y**. Ranking of the 11 variables by VIP values reveals their order of importance as follows: log time (1.400) > $J_{10\times10}$ (1.295) > $J_{5\times5}$ (1.185) > $J_{15\times15}$ (.985) > $J_{20\times20}$ (.865) > $J_{40\times40}$ (.847) > $J_{50\times50}$ (.843) > $J_{25\times25}$ (.8407) > $J_{30\times30}$ (.8406) > $J_{35\times35}$ (.838) > $J_{45\times45}$ (.835).

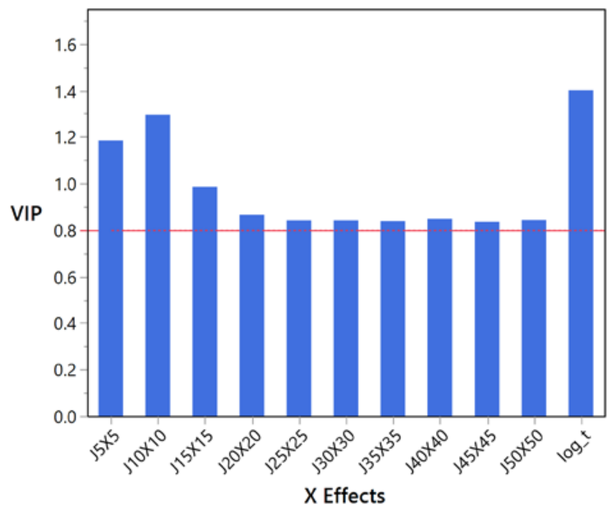

Figure 2: Variable Importance for Projection (VIP) for the Jaccard Coefficients

Despite being weaker than the top four variables (namely *log time*, $J_{10\times10}$, $J_{5\times5}$, and $J_{15\times15}$, in that order), the remaining Jaccard variables all have relatively large VIP scores. According to Cox & Gaudard (2013), only variables with small VIP scores and small regression coefficients should be considered for removal from the model (Wold, 1995). As the *Jaccard* coefficients all have large VIP scores, they should all be retained in the model, and not be pruned. Clearly, *log time,* $J_{10\times10}$, and $J_{5\times5}$ showed a much large influence on the factors than $J_{15\times15}$.

There appeared to be a negative correlation between *log time* and the *Jaccard* variables, as shown in Figure 3. This observation agrees with the correlational analysis result that (log) time was inversely correlated with *Jaccard* measures (Table 1). It is not difficult to understand the phenomenon as novices who are at a greater 'distance' from the experts (smaller similarity indices) also tend to take much longer time in finishing the game.

The X Loadings plot (Figure 4) shows that not all predictors impact the factors in the same way. Further confirmation can be seen in the X Loadings Scatterplot Matrix (Figure 5).

Figure 5 shows the X Loadings Scatterplots Matrix for Factor 1. The plot revealed the *Jaccard* variables (right) to load differently from *log time* (left) for Factor 1, which clearly separates the similarity indices from the *time* function.

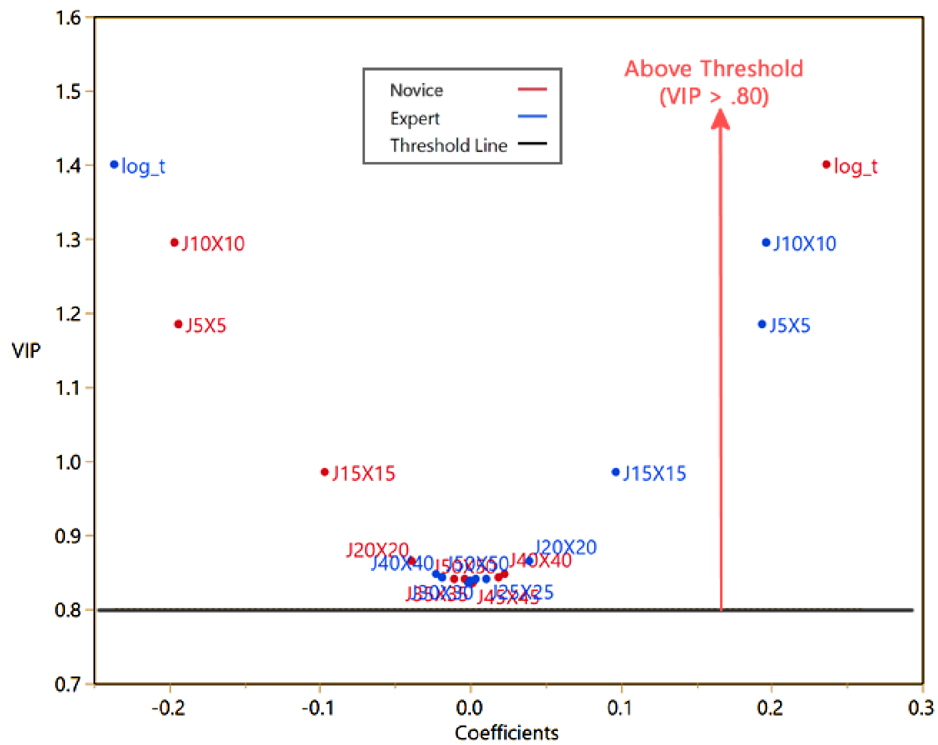

Figure 3: VIP vs Coefficients for Central and Scaled Data

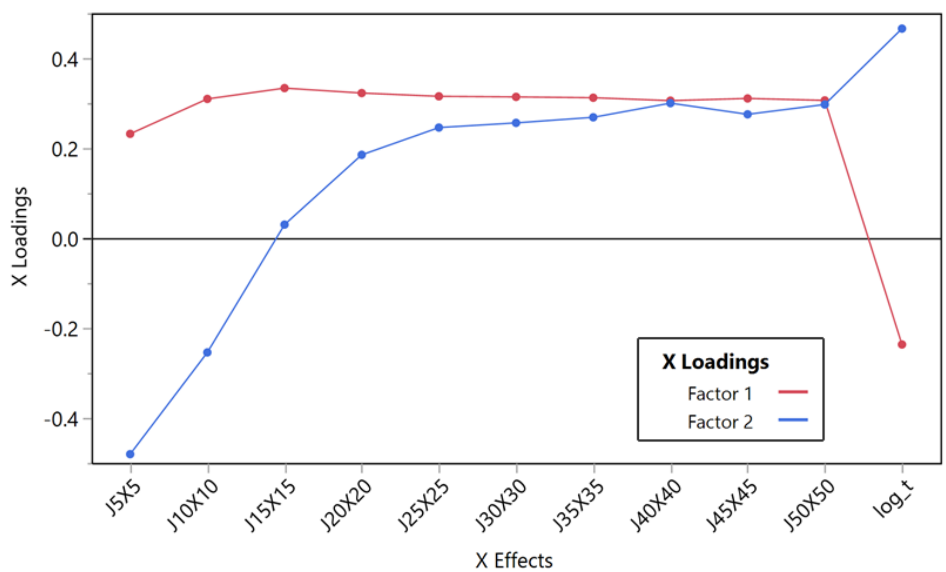

Figure 4: The X Loading Plots

Figure 6 shows the X Loadings Scatterplot Matrix for Factor 2 and tells a different story. The plot reveals $J_{10\times10}$ and $J_{5\times5}$ (bottom) to load differently from the *log time* and the remaining *Jaccard* coefficients for finer grids (from $J_{20\times20}$ to $J_{50\times50}$) (top). Interestingly, $J_{15\times15}$ only has a small effect on Factor 2 – indicated by its position near 0.

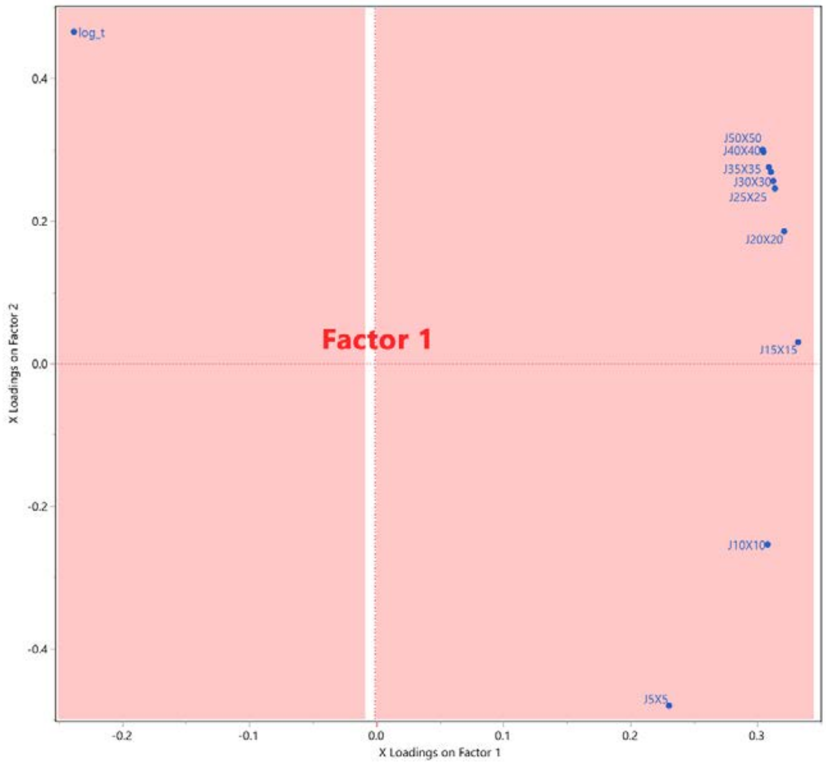

Figure 5: X Loadings Scatterplot Matrix for Factor 1

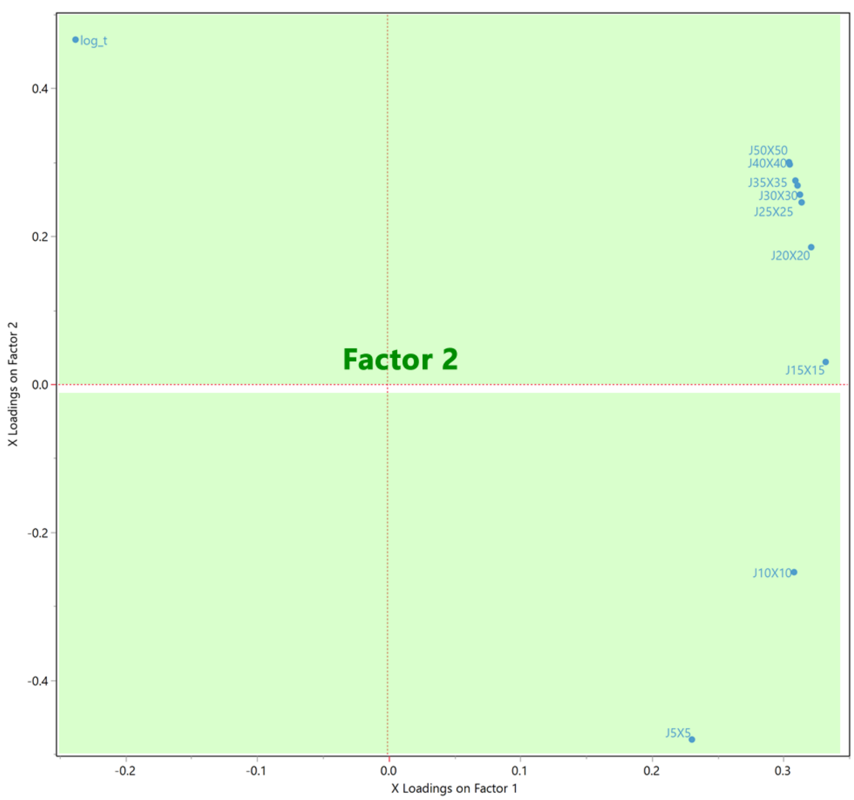

Figure 6: X Loadings Scatterplot Matrix for Factor 2

Discussions

Commercial games commonly relied on *time* – i.e., the ability of players to meet objectives in the shortest possible time, for the calculation of high scores on Leaderboards, thus, serving as a key indicator for superior performance. While time is indeed a very good indicator, it is not the only one that undergird performance. Over-reliance on time as a Key Performance Indicator (KPI) risked ignoring players' decisions in other important matters, such as path selections, order of mission completion, strategizing, etc. – many of which are much more relevant to efficiency in mission accomplishment (particularly for military operations). Procedural orders of events often impact learning/training performance because many tasks and learning objectives are, in fact, hierarchical in nature (Jonassen, Hannum, & Tessmer, 1989).

When compared to entertainment games, time is often not a good KPI for *serious games* (Loh & Sheng, 2014) because many training and learning situations require the learners to complete objectives in a certain order, stressing strategies over speed. Time pressure not only put workers/learners under duress but also lead them into risky behaviors: e.g., taking unnecessary chances, making hasty decisions (e.g., Ben Zur & Breznitz, 1981; Pieters & Warlop, 1999; Young, Sutherland, & Cole, 2011), even workplace disasters (e.g., Wickens, Stokes, Barnett, & Hyman, 1993). Game players who are already familiar with taking unrealistic risks in games due to the presence of 'saved games' are particularly prone to such at-risk behaviors.

Our findings (Figure 6) seem to indicate that the size of the game grids matters – i.e., coarser grids loads differently than finer ones. We believe this demarcation is an artifact of the signal-to-noise ratio. As the game grids become finer, more and more information is being generated or captured. However, having more information does not necessarily mean better signals because some of the information gathered are simply interference or noise. Researchers should exercise caution and practice parsimony. There needs to be a balance in the signal-to-noise ratio, and the sweet spot can be easily determined by scanning the grid landscape to identify the right size of grid for investigation.

Conclusions

In this study, we used PLS DA as a classification method to predict expert-novice performance for the purpose of serious games analytics. By using PLS DA, we reduced the dimensions from ten predictors to a two-factor model (parsimony) that still accurately predicts as many experts/novices as possible in our data set.

Similarity indices and action sequences, from which similarity indices were derived, deserve a closer look from the serious games community because they can reveal hidden patterns about players' performance, not explainable by game playtime alone. Dividing the game map into grids of different sizes and analyzing players' movement, actions, and behaviors within these grids is but *one* way to obtain new features for serious games analytics.

This study shows that coarser grids may not yield enough information, while too fine a grid could result in too much information (and noise), thus becoming counter-productive and computationally wasteful. The optimum granularity to divide game world into grids is unique to each and every serious game produced – i.e., the optimal grid size is dependent on the way a particular game was designed (e.g., placement of events, navigational paths, etc). Researchers and analysts should perform a quick scan to identify the correct grid size for the serious game under investigation, so as to obtain the optimum amount of information for serious games analytics.

Proving the Efficacy of Serious Games

Given the correct KPIs, it would be easier for Chief Learning Officers of learning organizations to calculate Return of Investment, justify the bottom-line, and recommend the purchasing of serious games for training in the future. Action sequences of game-players deserve closer examination in serious games research because it reflects the cognitive-behavioral activities of game players and can contribute to the knowledgebase of efficacy in game-based learning or trainings. More research in serious games analytics will make the differentiation of experts from novices and the predictive identification of likely-experts from massive groups of trainees much easier.

The research at the Virtual Environment Lab has already identified a number of KPI for serious games analytics, including (a) Similarity Index for action sequences under a single expert, (b) Maximum Similarity Indices (MSI) when multiple experts are present, (c) objective- and navigational-based action sequences, and (d) the Expertise Index. When combined, three of them – particularly the Objective action sequences, Navigational action sequences, and Time have proven to be extremely useful for the measurement of Performance Improvement Index.

ACKNOWLEDGMENTS

This research was made possible in part through funding from the Defense University Research Instrumentation Program (DURIP) from the U.S. Army Research Office. The authors wish to thank Dr. Yanyan Sheng for her contributions in an earlier draft of this paper, Mr. Ting Zhou and Dr. JaeHwan Byun for their assistance in data collection, and Ms. Ariel Yining Loh for editing the manuscript.

REFERENCES

Alvarez, J., Djaouti, D., Rampnoux, O., & Alvarez, V. (2011). Serious games market: Some key figures (from 1950's to 2000's). Retrieved from http://serious.gameclassification.com/files/articles/sgc_report_03-11.pdf

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods*, *5*(16), 3790–3798. doi:10.1039/C3AY40582F

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of Serious Games: An overview. *Advances in Human-Computer Interaction*, *2013*, 11. doi:10.1155/2013/136864

Boulesteix, A.-L., & Strimmer, K. (2007). Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, *8*(1), 32–44. doi:10.1093/bib/bbl016

Cappiello, C., Matera, M., Picozzi, M., Sprega, G., Barbagallo, D., & Francalanci, C. (2011). DashMash: A mashup environment for end user development. In S. Auer, O. Díaz, & G. A. Papadopoulos (Eds.), *Web Engineering* (Vol. 6757, pp. 152–166). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-22233-7

Charles Krulak. (1997). *Military thinking and decision making exercises* (No. 1500.55). Washington, DC. Retrieved from www.marines.mil/Portals/59/Publications/MCO 1500.55.pdf

Cox, I., & Gaudard, M. (2013). *Discovering Partial Least Squares with JMP*. Cary, NC: SAS Institute.

Crookall, D. (2010). Serious games, debriefing, and simulation/gaming as a discipline. *Simulation & Gaming*, *41*(6), 898–920. doi:10.1177/1046878110390784

Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard Business Review*, *85*(7-8), 114–21, 193. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17642130

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, *64*(2), 79–102. doi:10.1016/j.ijhcs.2005.06.002

Jonassen, D. H., Hannum, W. H., & Tessmer, M. (1989). *Handbook of Task Analysis Procedures*. Westport, CT: Praeger Publishers.

Kozma, R. B., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching*, *34*(9), 949–968.

Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (2004). Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications (ETRA '04)* (pp. 41–48). New York, NY: ACM Press. doi:10.1145/968363.968370

Loh, C. S. (2011). Using in situ data collection to improve the impact and return of investment of game-based learning. In *Proceedings of the 61st International Council for Educational Media and the XIII International Symposium on Computers in Education (ICEM&SIIE'2011) Joint Conference* (pp. 801–811). Aveiro, Portugal. Retrieved from http://www.icem-cime.org/images/stories/icemsiie2011_Proceedings.pdf

Loh, C. S. (2012a). Improving the impact and return of investment of game-based learning. *International Journal of Virtual and Personal Learning Environments*, *4*(1), 1–15. doi:10.4018/jvple.2013010101

Loh, C. S. (2012b). Information Trails: In-process assessment of game-based learning. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in Game-based Learning: Foundations, Innovations, and Perspectives* (pp. 123–144). New York, NY: Springer. doi:10.1007/978-1-4614-3546-4

Loh, C. S., Anantachai, A., Byun, J. H., & Lenox, J. (2007). Assessing what players learned in serious games: In situ data collection, Information Trails, and quantitative analysis. In Q. Mehdi (Ed.), *Proceedings of the Computer Games: AI, Animation, Mobile, Educational & Serious Games Conference (CGAMES 2007)* (pp. 10–19). Wolverhampton, UK: University of Wolverhampton.

Loh, C. S., & Sheng, Y. (2013). Performance metrics for serious games: Will the (real) expert please step forward? In *Proceedings of CGAMES'2013 USA* (pp. 202–206). IEEE. doi:10.1109/CGames.2013.6632633

Loh, C. S., & Sheng, Y. (2014). Maximum Similarity Index (MSI): A metric to differentiate the performance of novices vs. multiple-experts in serious games. *Computer in Human Behavior*, *39*, 322–330. doi:10.1016/j.chb.2014.07.022

Loh, C. S., & Sheng, Y. (2015a). Assessing expert performance as Serious Games Analytics: From data to actionable insights. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*. New York, NY: Springer.

Loh, C. S., & Sheng, Y. (2015b). Measuring the (dis-)similarity between expert and novice behaviors as serious games analytics. *Education and Information Technologies*, *20*(1), 5–19. doi:10.1007/s10639-013-9263-y

Loh, C. S., Sheng, Y., & Ifenthaler, D. (Eds.). (2015). *Serious Games Analytics: Methodologies for Human Performance Measurement, Assessment, and Improvement*. New York, NY: Springer.

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious games analytics: Theoretical framework. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious Games Analytics: Methodologies for Human Performance Measurement, Assessment, and Improvement*. New York, NY: Springer.

Loh, C. S., Sheng, Y., & Li, I.-H. (2015). Predicting expert-novice performance as Serious Games Analytics: Objective-oriented vs. navigational (spatial-temporal) action sequences. *Manuscript Submitted for Publication*.

Mishra, J., Zinni, M., Bavelier, D., & Hillyard, S. A. (2011). Neural basis of superior performance of action videogame players in an attention-demanding task. *The Journal of Neuroscience*, *31*(3), 992–8. doi:10.1523/JNEUROSCI.4834-10.2011

Monge, A. E., & Elkan, C. P. (1997). An efficient domain-independent algorithm for detecting approximately duplicate database records. In *Proceedings of the SIGMOD 1997 Workshop on Research Issues on Data Mining and Knowledge Discovery* (pp. 23–29). Tuscon, AZ.

Thawonmas, R., Yoshida, K., Lou, J.-K., & Chen, K.-T. (2009). Analysis of area revisitation patterns in World of Warcarft. In S. Natkin & J. Dupire (Eds.), *8th International Conference of Entertainment Computing (ICEC 2009)* (Vol. 5709, pp. 13–23). Berlin/Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-04052-8

Thawonmas, R., Yoshida, K., Lou, J.-K., & Chen, K.-T. (2011). Analysis of revisitations in online games. *Entertainment Computing*, *2*(4), 215–221. doi:10.1016/j.entcom.2011.01.003

Winkler, W. E. (2006). *Overview of Record Linkage and current research directions*. Washington, DC. Retrieved from http://www.census.gov/srd/papers/pdf/rrs2006-02.pdf

Wold, S. (1995). PLS for multivariate linear modeling. *Chemometric Methods in Molecular Design*, *2*, 195.

Authors' Biographies

Christian Sebastian Loh, Ph.D. (Corresponding Author)

Director, Virtual Environment Lab (V-LAB)
Department of Curriculum & Instruction
Southern Illinois University Carbondale
625 Wham Drive, Carbondale, IL 62901-4610, U.S.A.
csloh@siu.edu

Dr. Loh is Director of the Virtual Environment Lab at the Southern Illinois University Carbondale and an awardee of the Defense University Research Instrument Program (DURIP) grant. His research and professional interests include serious games for training, performance differences between experts and novices, decision-making in virtual environment, and information visualization for serious games analytics. He is the lead editor of a new book by Springer, entitled: *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*.

I-Hung Li, M.S.

Virtual Environment Lab (V-LAB)
Department of Curriculum & Instruction
Southern Illinois University Carbondale
625 Wham Drive, Carbondale, IL 62901-4610, U.S.A.
henryryu@siu.edu

Mr. I-Hung Li is a software engineer and doctoral candidate in the Learning Systems Design and Technology (LSDT) program. He holds a Master's degree in Electrical and Computer Engineering and is an assistant researcher in the Virtual Environment Lab (V-Lab) at the Southern Illinois University Carbondale. His research interests include performance assessment in virtual environments, performance data visualization, and incorporating engineering algorithms into state-of-the-art game engines for analytics production. He is currently working on integrating the Unity3D game engine with *Performance Tracing Report Assistant* (PeTRA) in the *Information Trails* assessment framework for Serious Games Analytics.